An Annotated Catalog of Salivary Gland Transcripts from *Ixodes scapularis* Ticks

José M. C. Ribeiro [1,*], Francisco Alarcon-Chaidez[2], Ivo M. B. Francischetti [1], Ben J. Mans [1], Thomas N. Mather [3], Jesus G. Valenzuela [1], and Stephen K. Wikel [2]

[1] Section of Vector Biology, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, United States of America, [2] Center for Microbial Pathogenesis, School of Medicine, University of Connecticut Health Center, Farmington, Connecticut, United States of America, [3] Center for Vector-Borne Disease, University of Rhode Island, Kingston, Rhode Island, United States of America

- To whom correspondence should be addressed.

Jose M.C. Ribeiro

NIAID/LMVR

12735 Twinbrook Parkway room 2E32D

Rockville MD 20852

Phone: 301-496-9389

E-mail: jribeiro@niaid.nih.gov

Running title: Updated Sialome of *Ixodes scapularis*

Key Words: Transcriptome; salivary glands; sialome; hematophagy; gene duplication.

Abbreviations: AA, amino acid; EST, expressed sequence tag; H, housekeeping category of expressed genes; HBP, histamine-binding protein; NCBI, National Center of Biological Information; ORF, open reading frames; S, putatively secreted category of expressed genes; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; TIL, trypsin inhibitor-like; U, unknown category of expressed genes.

Abstract

Over 8,000 expressed sequence tags from six different salivary gland cDNA libraries from the tick *Ixodes scapularis* were analyzed. These libraries derive from feeding nymphs infected or not with the Lyme disease agent, *Borrelia burgdorferi*, from unfed adults, and from adults feeding on a rabbit for 6–12 hours, 18–24 hours, and 3–4 days. Comparisons of the several libraries led to identification of several significantly differentially expressed transcripts. Additionally, over 500 new predicted protein sequences are described, including several novel gene families unique to ticks; no function can be presently ascribed to most of these novel families. Among the housekeeping-associated transcripts, we highlight those enzymes associated with post translation modification of amino acids, particularly those forming sulfotyrosine, hydroxyproline, and carboxyl-glutamic acid. Results support the hypothesis that gene duplication, most possibly including genome duplications, is a major player in tick evolution.

Supplemental spreadsheets with hyperlinks to all sequences used in this manuscript can be found at http://www.ncbi.nlm.nih.gov/projects/omes/#salivary_transcriptomes, and are hyperlinked throughout the text. Static original data is also presented as a set of compressed files at the publisher's site at http:/www.####################################

Introduction

Ticks are highly efficient arthropod vectors of infectious diseases, being able to transmit viruses, bacteria, protozoa, and helminth parasites. The tick *Ixodes scapularis* (= *Ixodes dammini*) is a vector for such human diseases as Lyme borreliosis, human granulocytic anaplasmosis (= ehrlichiosis), and babesiosis in the Central and Eastern United States. *I. scapularis* belongs to the subarctic *Ixodes persulcatus* complex of species that include *Ixodes ricinus* and *Ixodes pacificus,* vectors of Lyme disease in Asia, Europe and the western U.S., respectively. *I. scapularis* has three life stages after hatching from the egg: Larvae and nymphs feed normally on small rodents, while adults feed on medium to large mammals (Anderson, 2002).

After finding a suitable site to feed, hard ticks penetrate the host skin with their mandibles and secrete a glue-like substance that helps to anchor and prevent detachment of the arthropod from the feeding site. Feeding may be accomplished in days or weeks, by the end of which the tick usually has increased in weight by up to 1,000 fold its initial size (Kaufman, 1989). Feeding is accomplished by sucking blood from a hematoma, or pool of blood, formed in the host skin by mouthpart laceration of the host tissues. Maintenance of the liquid state of this pool is accomplished by a complex mixture of antihemostatic compounds present in tick saliva (Ribeiro, 1989; Ribeiro, 1995). Indeed, ticks alternately feed and salivate in the host skin (Gregson, 1967). Tick saliva also contains antiinflammatory and immunomodulatory compounds that prevent immune reactions from disrupting the feeding process (Wikel, 1999; Wikel and Alarcon-Chaidez, 2001). Probably due to these pharmacologic activities, saliva enhances pathogen transmission, and antisaliva immunity affects pathogen transmission (Gillespie et al., 2000). Accordingly, knowledge of salivary components in vector ticks can lead to the discovery

of novel pharmacologic molecules and to the development of novel transmission-blocking vaccine targets against the diseases that ticks transmit.

Among the pharmacologically active compounds in the saliva of *I. scapularis* are prostaglandins (Ribeiro et al., 1985; Ribeiro et al., 1988), an anticomplement named ISAC (Ribeiro et al., 1985; Ribeiro et al., 1988), several anticlotting proteins (Francischetti et al., 2004; Francischetti et al., 2002a; Narasimhan et al., 2002), a fibrin(ogen)olytic metalloprotease(Francischetti et al., 2003), an immunosuppressive protein (Anguita et al., 2002), and undefined molecules with kininase (Ribeiro and Mather, 1998), antineutrophil (Ribeiro et al., 1990) and anti-IL-2 (Gillespie et al., 2001) activities. Many putative proteins and peptides, including some of those described above (Francischetti et al., 2003; Francischetti et al., 2004; Francischetti et al., 2002a), were discovered following a sialotranscriptome analysis of *I. scapularis*, based on 735 expressed sequence tags (EST) (Valenzuela et al., 2002). In the present work, we extend the analysis of the sialotranscriptome of *I. scapularis* to include transcripts from libraries made not only at 3–4 days after attachment of adult female ticks to their hosts, as done before (Valenzuela et al., 2002), but also including libraries constructed from female adult ticks that were unfed, fed for 6–12 hours, fed for 18–24 hours, and from nymphs feeding 24–48 hours. At least 1,000 EST were sequenced per library, for a total of 8,150 EST. Many genes were significantly differentially represented in these libraries. Nearly 500 new full-length sequences are described, including both housekeeping and putatively secreted gene products. Among the housekeeping proteins, the finding of proline hydroxylases and sulfotransferases suggests that secreted products may be modified as collagen is or may contain sulfotyrosine residues. Among the secreted proteins, it appears that most belong to lineage-specific gene family expansions, some of which appear to exhibit high levels of polymorphism for particular genes/proteins, providing a large repertoire of pharmacologically active components.

Methods

Ticks. Pathogen-free adult female *I. scapularis* ticks were obtained either from colonies maintained at the University of Connecticut Health Center (6–12 and 18–24 libraries) or field-collected specimens from the Center for Vector-borne Disease at the University of Rhode Island) (all other libraries). Adult ticks were fed on rabbits. Nymphs were fed on hamsters previously infected or non-infected with *Borrelia burgdorferi.* Salivary glands were dissected from either 20–30 female ticks or from 100 nymphs per library. Glands were briefly washed in ice-cold 1x PBS and immediately stored in RNAlater storage solution (Ambion, Austin, TX USA) or in phosphate buffered saline at –80°C until used for library synthesis.

Tissue preparation and cDNA library synthesis. Poly A^+ mRNA was isolated from tick salivary glands with the Oligotex direct mRNA microkit from Qiagen (Chatsworth, CA USA), or the micro Fast track mRNA isolation kit (Invitrogen, Carlsbad, California) following the manufacturer's protocol. cDNA libraries were constructed in the vector pTriplEx2 from this RNA using the SMART cDNA library construction kit (BD Clontech, Palo Alto, Ca USA). Random clones were sequenced from the 5' direction only, because successful sequencing from the 3' end was usually lower than 40%. Full length sequences were obtained in selected cases by performing primer-based extension protocols. Consult Francischetti et al. (Francischetti et al., 2002b) for more details.

Bioinformatic tools used. EST were trimmed of primer and vector sequences, clusterized, and compared with other databases as described before (Valenzuela et al., 2003). The BLAST tool (Altschul and Gish, 1996) and CAP3 assembler (Huang and Madan, 1999) were used for the clusterization. First, sequences that had <5% N were blasted against each other

and the output scanned to join in a single fasta file those that had minimum identity of 81 AA over a 90-AA stretch. These sequences were, in turn, given as input to the CAP3 assembler, which could further divide the blast cluster into additional clusters. The final output was piped into a tab-delimited file imported into an Excel (Microsoft Excel Analysis Tools, Seattle, WA.USA) spreadsheet. These operations were automated by a program written in Visual Basic named Cluster5. O-glycosylation sites on the predicted proteins were obtained with the program NetOGlyc (http://www.cbs.dtu.dk/services/NetOGlyc/) (Hansen et al., 1998). We submitted all translated sequences (starting with a Met) to the Signal P server (Nielsen et al., 1997) to detect signal peptides indicative of secretion. BLAST searches were done locally from executables obtained at the NCBI FTP site (ftp://ftp.ncbi.nih.gov/blast/executables/) (Altschul et al., 1997) against the non-redundant protein database of the NCBI, the gene ontology fasta subset (Lewis et al., 2000), the conserved domains database of NCBI (Marchler-Bauer et al., 2002) containing the KOG (Tatusov et al., 2003), Pfam (Bateman et al., 2000) and Smart (Schultz et al., 2000) motifs and to custom-downloaded databases containing mitochondrial and rRNA nucleotide sequences available at the NCBI. Phylogenetic analysis were done using the program ClustalX (Thompson et al., 1997) and the MEGA3 package when there were too many divergent sequences (Kumar et al., 2004). Phylogenetic trees were constructed with the neighbor-joining algorithm and were bootstrapped 10,000 times. Gaps were taken into consideration by the pairwise deletion method. AA substitutions were considered by a Poisson model. The values shown on the nodes of the trees represent the percentage of times there was support for the branching on the 10,000 trials. The names of the sequences in the alignments and phylograms consists of either a five-letter abbreviation of the species name (the first three from the genus followed by the first two of the species name, thus IXOSC for *Ixodes scapularis*) followed by the NCBI identifier number that follows the gi| for sequences deposited at GenBank at the time of

manuscript writing; otherwise, the sequences reflect the arbitrary names of the clones or clusters of clones that were sequenced in this work.

Supplemental material: Supplemental tables I and II and figures are hyperlinked throughout the paper to the NCBI page http://www.ncbi.nlm.nih.gov/projects/omes/ where this and other transcriptomes are located. For record purposes, zipped files of the supplemental tables 1 and 2 can also be found as supplemental material at Elsevier. The archives for each supplemental data set (6 files for supplemental table 1 and 5 for supplemental table 2) should be downloaded and extracted to 2 newly created directories in the user's computer, and each Excel file should then be open for browsing.

Results and Discussion

cDNA Library Characteristics

Six different cDNA libraries from *I scapularis* nymphs and adult female ticks were used in this work, as follows: 1) nymphs fed for 2 days, uninfected (NUNF); 2) nymphs fed for 2 days, previously fed as larvae on *Borrelia burgdorferi*-infected mice (NINF); 3) adult females, unfed (AUF); 4) adult females, 6–12 hours post host attachment (6–12); 5) adult females 18-24 hours post host attachment (18–24); and 6) adult females 3–4 days post host attachment (72). A total of 8,150 clones were sequenced in the combined libraries (Table 1) including 735 sequences previously reported (Valenzuela et al., 2002). Supplemental Table 1 contains links to all EST used in the present work.

Description of the Clusterized Data Set

Of the 8,150 clones sequenced, 7,476 were considered of high quality—having less than 5% indeterminate base calls—and were used to assemble a clusterized database (Supplemental

Table 1), yielding 3,020 clusters of related sequences, 2,333 of which contained only one EST. The consensus sequence of each cluster is named either a contig (deriving from two or more sequences) or a singleton (deriving from a single sequence); in this paper, for simplicity sake, we will use the denomination contig to address sequences deriving both from consensus sequences and from singletons. The 3,020 contigs were compared by the program blastx, blastn, or RPSBLAST (Altschul et al., 1997) to the nonredundant protein database of the National Center of Biological Information (NCBI), to the gene ontology database (Ashburner et al., 2000), to the conserved domains database of the NCBI (Marchler-Bauer et al., 2002),and to a custom-prepared subset of the NCBI nucleotide database containing either mitochondrial or rRNA sequences. Because the libraries used are unidirectional, the three frame translations of the dataset were also derived, and open reading frames (ORF) starting with a methionine and longer than 40 amino acid (AA) residues were submitted to SignalP server (Nielsen et al., 1997) to help identify putative secreted proteins. The EST assembly, BLAST, and signal peptide results were piped into an Excel spreadsheet for manual annotation.

Six categories of expressed genes derived from the manual annotation of the contigs (Table 2). The putatively secreted (S) category contained 29% of the clusters and 49% of the sequences, with an average number of 4.2 sequences per cluster. The housekeeping (H) category had 37% and 36% of the clusters and sequences, respectively, and an average of 2.45 sequences per cluster. Thirty-four percent of the clusters, containing 15% of all sequences, were classified as unknown (U) because no assignment for their function could be made; most of these consisted of singletons. A good proportion of these transcripts could have derived from 3′ or 5′ untranslated regions of genes of the above two categories, as was recently indicated for a sialotranscriptome of *Anopheles gambiae* (Ribeiro, manuscript submitted). Possibly transposable elements originated 35 clusters, mostly singletons, suggesting there is active ongoing

transposition in *I. scapularis* ticks. *Babesia*-derived genes were found in three sequences, and a densovirus transcript also possibly reflected an infected tick or a laterally transferred gene.

Housekeeping (H) Genes

The 1,106 gene clusters (comprising 2,704 EST) attributed to H genes expressed in the salivary glands of *I. scapularis* were further characterized into 21 subgroups according to function (Table 3). Not surprisingly for an organ specialized in secreting polypeptides, the two larger sets were associated with protein synthesis machinery (862 EST in 179 clusters) and with energy metabolism (mostly mitochondrial enzymes; 154 clusters containing 732 EST). These two sets differed from the remaining 19 categories by having a larger average of EST per cluster, above 4.5, while all remaining clusters had 2.14 or lower. We have also included in this category a group of 239 EST that grouped into 163 clusters and represent conserved proteins of unknown function presumably associated with cellular metabolism. The group of clusters attributed to the protein modification function is of interest due to the specialized function of the salivary glands of ticks. This group, comprising 114 EST grouped into 60 clusters, includes chaperones, enzymes associated with disulfide bridge formation, and also proline hydroxylases subunits and sulfotransferases, indicating possible AA modifications that might occur in the mature secreted salivary polypeptides of *I. scapularis*. Additional inspection of each of the 1,106 gene clusters for further information in each can be done online with Supplemental Table 1.

Babesia-derived Transcripts

Three EST in the database produced blastx matches that strongly resemble apicomplexa sequences and could derive from field-collected adults infected with *Babesia microti.* Contig_128 produced a best match with gi|46226391, a protein from *Cryptosporidium parvum*.

Contig_1546 matches near perfectly gi|7716002, the seroreactive antigen BMN1-9 of *B microti,* and contig_2156 matches gi|46226717, a large low-complexity coiled-coil protein with a large repeat region from *C. parvum*. Several other proteins in the database produced best matches to Plasmodium proteins, but these were usually of low complexity and were not annotated as babesial.

Viral Transcripts

Contig_2698 produced a best blastx match with gi|9633607, a nonstructural protein from the *Periplaneta fuliginosa* densovirus, scoring significantly with several other reported densoviral proteins. Densovirus belongs to the Parvoviridae family; its members are known to infect insects (O'Neill et al., 1995), but no tick infections have been reported.

Differential Salivary Gland cDNA Library Representation of Expressed Genes

Following clusterization of the combined data for the six libraries, we observed that some clusters of related sequences contained more or fewer sequences from a particular value than expected from a random distribution, as evaluated by the χ^2 test. We found differential gene expression between infected and noninfected nymphs, between nymphs and adults, and between adults at different stages of feeding, as follows.

Differential Salivary Gene Expression Between *B. burgdorferi*-infected (NINF) and Noninfected Nymphs (NUNF)

Only 10 EST clusters were significantly differentially represented when the number of EST from each of the two libraries was compared in each of the sequence clusters. In all cases, these clusters represented translation products associated with secreted proteins. Of the

10 clusters, 7 were overrepresented in the NINF library and 3 were overrepresented in the NUNF library. The clusters containing overrepresented sequences from the NINF library (Table 4A), thus possibly associated with an immune response, coded for products of the 5.3-kDa peptide family (4 clusters), two overrepresented clusters of the basic tail family (Valenzuela et al., 2002), and one cluster coding for a histamine-binding protein (HBP), a large family of salivary proteins that may act by binding to pharmacologically active agonists (Paesen et al., 1999). On the other hand, all three overrepresented sequences in the NUNF library were of the HBP family (Table 4B). While further studies are needed to explore these findings, the data are suggestive that the 5.3-kDa gene family, which has no known function, may be implicated in defense responses. The relatively small number of differentially translated gene products following infection by *B. burgdorferi* is paralleled by a work done recently with the tick *Rhipicephalus appendiculatus* salivary glands infected and not infected with the protozoan parasite *Theileria parva* (Nene et al., 2004), which actually found no significantly differentially expressed genes between the two randomly sequenced libraries, with a data set totaling over 18,000 EST.

Differential Salivary Gene Expression Between Adults and Nymphs

To compare the gene expression in the salivary glands of nymphal ticks to that of adult ticks, we pooled together the data derived from the nymphal libraries, because they were mostly not significantly different from each other, and to increase the sensitivity of the χ^2 test. Accordingly, we observed 7 clusters that had overrepresentation of nymphal EST (Table 5A), including the previously described salp9 polypeptide (most probably a truncated form of an anti-complement protein – see below), which was identified from a salivary nymphal cDNA library (Narasimhan et al., 2002) and not previously found in an analysis of 765 EST from the salivary glands of adult female ticks (Valenzuela et al., 2002), 4 clusters from the basic tail family

(thought to be mostly anticlotting peptides—see below), one member of the 5.3-kDa family (of unknown function) that was included in the overrepresented NINF library (contig-123) (Table 4A), and one HBP (possible binder of pharmacologic agonists). Although these differences could derive from differential time expression of genes, they otherwise could indicate that there are some genes that are predominantly expressed in nymphs as opposed to adults.

Table 5B indicated clusters where nymphal sequences were underrepresented, including 5 clusters coding for products with high proline and glycine, and thus collagen-like, which might function as dermal glue or tick attachment peptides; in addition to other putative secreted proteins including Kunitz domain-containing proteins indicative of serine protease inhibitors, other cysteine-rich peptides, and a metalloprotease. The clusters indicated in Table 5B are probably artifactual due to the more complete time-of-feeding representation of the adult libraries compared with the nymphal libraries, which were constructed from a single time point. Indeed, the collagen-like peptides are more represented on the adult libraries constructed at 6-12 and 18–24 hours post host attachment, as are some of the other transcripts.

Differential Salivary Gene Expression in Adult Female Ticks as a Function of Time Post Host Attachment

The four salivary gland cDNA libraries from adult female ticks feeding for different intervals of time were compared to identify significantly overexpressed transcripts. In the AUF library, only one gene was found overexpressed (contig 145—see Supplemental Table 1), coding for an 18S rRNA. This possibly reflects the readiness of the gland to initiate protein synthesis after host attachment and dilution of this transcript after transcriptional activation of the salivary gland following attachment, but may also be an artifact of mRNA extraction, since only polyadenylated transcripts should be cloned with the library construction method used. The

library made with ticks attached for 6–12 hours to a host showed no significantly increased transcripts; however, the library constructed with ticks attached for 18–24 hours post host attachment showed 20 gene products 2- or more fold significantly overrepresented, 19 of which coded for putative secreted proteins and one coding for an elongation factor (Table 6). Five of the overexpressed clusters coded for collagen-like peptides, four for basic tail products, two for Kunitz domain-containing peptides, and the remaining for other products of unknown function. Finally, the library made from ticks feeding for 3-4 days contained transcripts significantly overrepresented in 7 clusters of putative secreted peptides, each of a different protein family (Table 7).

Although the results comparing transcript abundance between the six libraries are relatively modest in the number of gene products identified as differentially transcribed, they suggest *i)* that the 5.3 kDa peptide family may be involved in immune responses to bacteria; *ii)* strong evidence for differential gene expression, of at least a few gene products, between nymphal and adult ticks; and *iii)* evidence for differential gene expression in the salivary glands of adult ticks as a function of time post attachment, with different members of the same gene family being expressed preferentially at different times. Microarray experiments based on the transcripts provided in this database (Supplemental Table 1) would be a natural follow up for testing these hypotheses, particularly if single ticks could be used in the array experiments.

An Updated Catalog of *I. scapularis* Salivary Gland Transcripts

Several clusters indicated in Supplemental Table 1 are abundant enough to extract consensus sequences of novel sequences. Additionally, we have performed primer extension studies in several clones to obtain full- or near full-length sequences of products of interest. These novel sequences (511), plus the previously known salivary gland-derived sequences from

I. scapularis deposited at NCBI (143), are grouped together in Supplemental Table 2. Following our recent nomenclature of the salivary gland products of *I. pacificus* (Francischetti et al., 2005), the salivary gland products of *I. scapularis* are divided into 27 Groups as follows:

Group 1: basic tail polypeptides. Polypeptides of this family are recognized by having generally 13–14 kDa and one or more of the following patterns or motifs: aminoterminal region motif A-A-X(2)-C-X-N-G-T-R-P-X-S, a carboxyterminus motif F-F-X(3)-E-X-C-F-Y-X(2)-G-X(2)-G-X-C-X(2)-G-X-C-H-L-X(6)-P, a basic tail cysteine framework C-X(14)-C-X(3)-C-X(18)-C-X(9)-C, and either a carboxyterminus Lys-rich motif of type I, [EK]-P-K-Q-K-K-K-K-[TLPQ]-K-K-[TNPA]-K-K-P-K-R-[KN]-[ST]-K-K, type II, E-K-P-[RK]-K-[SN]-K-K-K-[SQ]-K-K-[ST]-K-K-P-K-K-[ST]-K-K-P, or type III, [PA]-[PSF]-K-K-[KE]-[KE]. Two members of this family are anomalously long, possessing inserts at different locations and producing a predicted protein of 18–19 kDa (Figure 1A). In both cases, these inserts are flanked by Gly or Pro, indicating that they form an extra loop in the protein. This family probably evolved by multiple gene duplication, as indicated by the multiple clades in the dendrogram (Figure 1B). Additionally, its members appear to be very polymorphic (Figure 1), as indicated by many sequences with small differences among themselves. The high polymorphism of this group is indicated by clusterization of the protein database to include relatives with 95% identity over at least 95% of the length (herein called 95-95 clusterization) (Supplemental Table 2); thus, 9 sequences are shown to be related to gi|22164172, 5 sequences are related to gi|22164156, 3 sequences each to gi|22164174 and gi|22164168, and 2 sequences to gi|22164184. Clusterization at 90% identity over 90% of protein length (90-90 clusterization) leads to identification of 8 protein clusters, suggesting at least 8 genes are related to basic tail proteins, or more if they reflect more recent gene duplication events. Edman degradation experiments from

salivary gland homogenates and saliva separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Valenzuela et al., 2002) identified many members of this family as being translated (Supplemental Table 2); however, the precise family member(s) expressed is still unknown due, to the limited size of the aminoterminal sequence obtained and due to the overlap in sequence identity of the many members of the family.

The function of this group of proteins is probably related to blood anticlotting activity. The protein coded by gi|15428308, named salp14, is the only basic tail family member that has been expressed and shown to have anticlotting activity, inhibiting factor Xa (Narasimhan et al., 2002). The cladogram in Figure 1B indicates 3 other closely related sequences and an *I. pacificus* clade also closely related to salp14. Other members of this family are also found in *I. pacificus* (Figure 1), where 2 clades are distinct in the dendrogram (Figure 1B). One additional *I. pacificus* sequence clusters with one *I. scapularis* clade, indicating that multiple members of the family probably existed before these two species separated. No sequence of this family is reported for *I. ricinus,* the European member of the *I. persulcatus* group of species to which *I. scapularis* and *I. pacificus* belong. As indicated in our recent study of *I. pacificus* sialotranscriptome, the basic tail of this family may help interaction of the protein with anionic phospholipids that are an essential component of the Xase and prothrombinase complexes (Andersen et al., 2004).

Group 2: basic tailless peptides. Peptides of this family are 10–11 kDa in mass and contain 1 or more of the 3 first motifs indicated for Group 1 peptides, but are shorter and do not have the poly lysine tail. Their similarity to Group 1 proteins is evident on the alignment and cladogram of Figure 1. Similarly to the basic tail proteins, this peptide family probably evolved by gene duplication. Protein clusterization (90–90) identifies one group with 11 members and

2 other singletons, indicating that at least 3 genes are involved in expression of basic tailless peptides. On the other hand, 95–95 clusterization indicates 13 different groups.

The function(s) of this group of peptides remains unknown. Supplemental Table 2 includes in this group an anomalous sequence (Is-6-12-J-cluster-394) that has the Cys framework of the family, a smaller molecular weight, and nothing else in common with the other peptides; it is not shown in Figure 1. It should be cautioned that an artifact of library construction may plague this group of peptides because the possibility exists that this group also has a poly Lys tail, but constituted of several AAA repeats that could be primed by the polyT primer used in the library construction. The original primer sequence would, in turn, be replicated to display a polyA at the end of the final sequenced clone. If this is the case, Group 2 would be constituents of Group 1. Genomic information on this group of peptides should settle this question.

Group 3: Kunitz domain-containing proteins. Proteins of this group either contain a Kunitz domain, as indicated by the PFAM or SMART algorithms, or they are related in sequence similarity to proteins with Kunitz domains, even though they do not themselves have strong domain signature. They are further subdivided into 3 groups: those containing a single Kunitz domain, those containing 2, and those containing 5 domains; accordingly, they are named, respectively, the monolaris, bilaris, and penthalaris subgroups (Francischetti et al., 2005).

To investigate the evolutionary relationships of this large group, the 32 published sequences of *I. scapularis* containing Kunitz domains were grouped with 47 new sequences containing this domain plus 14 sequences that are deposited at NCBI from other ticks having the KU Smart signature and containing a signal peptide indicative of secretion. The alignment and phylogram of these 92 Kunitz-containing tick salivary polypeptides (Figure 2) show the diversity of this group. The bootstrapped phylogram shows some robust clades, suggesting that multiple gene duplication has occurred during tick irradiation. Accordingly, four monolaris clades are

observed, two clades form the bilaris group, and one clade contains the penthalaris group. The six sequences from the closely related *I. pacificus* distributes among the *I. scapularis* sequences, indicating these genes were already present before irradiation of these two species. Notably, the five sequences obtained from the hard ticks *I. ricinus*, *Amblyomma hebreum*, and *Boophilus microplus* group with the monolaris II group, suggesting this group to be the most closely related to an ancestral salivary Kunitz gene expressed in tick saliva, because *Amblyomma* and *Boophilus* are more distantly related metastriate ticks, while *Ixodes* are prostriate (Black and Roehrdanz, 1998; Shao et al., 2004). The two *Ornithodorus* sequences do not group robustly with any of the other 90 sequences, as expected for this more distantly related group of soft ticks.

The monolaris group, defined as peptides with a single Kunitz domain or similar in sequence to other peptides having a Kunitz signature (after RPSBLAST to the Smart database) and having molecular mass smaller than 12 kDa, was analyzed separately from the bilaris and penthalaris group (Supplemental Table 2 and Figure 3). The phylogram suggests evolution by extensive gene duplication. Components of this group may have evolved from related members at a fast pace. Kunitz domains contain typically six Cys residues that form three disulfide bridges, and their spacing may be much more conserved. Accordingly, we can also organize the family into conserved motifs found among sequences, as follows:

- salp10 monolaris subgroup I (named after a member of this family) all have motifs C-X(5)-C-X(15)-C-X(8)-C-X(11)-C-X(3)-C and L-X(3)-C-X(2)-P-X(2)-C-X-G-X(6)-Y-Y-X(4)-G-C (Figure 4). These two motifs are not found in any other sequence shown in Supplemental Table 2. The 12 sequences of this subgroup are distributed in 3 distinct clades, with robust bootstrap support. The preservation of the two motifs suggests their origin from a common ancestral gene, followed by rapid evolution.

- The monolaris II subgroup, with a cysteine framework C-X(8)-C-X(18)-C-X(5)-C-X(12)-C-X(3)-C accounts for 16 different sequences (Supplemental Table 2; Figure 5) and are possibly represented by at least 10 different polymorphic genes.
- The monolaris III subgroup, with a Cys framework C-X(8)-C-X(15)-C-X(7)-C-X(12)-C-X(3)-C consists of six proteins in three distinct clades, indicating at least three distinct genes at work (Figure 6).

Seven other different cysteine frameworks found in the monolaris subfamilies are indicated in Supplemental Table 2 and will not be further described. Taken together, it appears that at least 20 different genes are responsible for producing members of this family, many of them abundantly transcribed and polymorphic. These peptides may function as protease inhibitors of the clotting cascade, because the Kunitz domain is typically found in serine protease inhibitors (Broze et al., 1990). None so far, however, has been characterized as such.

The bilaris or Ixolaris group of double Kunitz-containing salivary proteins comprises at least four clades as indicated in the phylogram of Figure 7, which includes the previously described anticlotting Ixolaris (gi|15077002) (Francischetti et al., 2002a; Monteiro et al., 2004). Ixolaris plus two other related sequences form clade 2 in Figure 7B. Another distinctive clade, indicated with number 1 in Figure 7B, is formed by five protein sequences that have a Ser- and Thr-rich carboxyterminus. They have 15 to 23 putative O-galactosylation sites as predicted by the NetOGlyc server (Hansen et al., 1998). These glycosylations may block protein degradation by carboxypeptidases in addition to targeting these proteins to yet-unknown host proteins or glycosaminoglycans. Only Ixolaris, among these 11 proteins, has been characterized and shown to inhibit the tissue factor pathway of blood clotting (Francischetti et al., 2002a; Monteiro et al., 2004), by interacting with the extrinsic Xase complex, which is a multiprotein complex assembled in a phospholipid surface. The binding of the dual Kunitz inhibitor to one member of

the proteolytic complex limits the diffusional loss of the inhibitor to find the second Kunitz target; thus, such dual-head inhibitors have much smaller effective concentrations for reaction inhibition than single-Kunitz proteins. These salivary dual-Kunitz proteins, accordingly, may target the intrinsic Xase and protrombinase complexes in addition to the extrinsic Xase complex.

The penthalaris group of salivary Kunitz-containing proteins groups itself into at least three clades (Figure 8). The eight sequences could represent the product of six genes if we consider some of the sequences to be alleles, although their divergence is higher than 10%. One member of this group, gi|22164266, has been expressed, and its activity is similar to that of Ixolaris, inhibiting the tissue factor pathway of blood clotting (Francischetti et al., 2004). These proteins may function in a similar redundant way to the members of the Ixolaris group.

No Edman degradation product was found in SDS-PAGE protein bands from saliva or salivary gland homogenates for any of the proteins containing Kunitz domains, either because these polypeptides are blocked or due perhaps to their low concentration, since they may be effective at Pico molar concentrations.

Group 4: proline- and glycine-rich peptides. This group comprises peptides usually containing 30% of their composition in AA Pro and Gly and is of predicted molecular mass ranging from 6 to 8 kDa. This family is also present in *I. pacificus*. Alignment of 60 sequences from *I. pacificus* and *I. scapularis* indicates a common ancestry of an already redundant family. Three clades are apparent for this group of peptides, which may derive from at least 10 different loci (Figure 9). Several of these peptides have the motif G-[SKQN]-[ST]-X-P-C-X(3)-P-G-X(2)-C-X(8)-P, as indicated in Supplemental Table 2. None of the members of this group has been characterized, and their function can only be speculated. Many of these peptides present weak similarity to collagens, containing a protein rich in Pro and hydroxylated proline through the post translation modification caused by the enzyme proline hydroxylase. Except for one

member of this group, all peptides shown in Figure 9 have one to six [LPAE]-P-G motifs that are known to be the targets of proline hydroxylase (Kivirikko et al., 1972; Rhoads and Udenfriend, 1969), and many posses the sequence Gly-Pro-Pro, which is known to interact with the platelet collagen receptor (Knight et al., 1999). Accordingly, these peptides may function as platelet aggregation inhibitors or tick attachment glue.

Group 5: similar to I. scapularis *18.7-kDa protein.* A protein named putative 18.7-kDa secreted protein was previously reported from *I. scapularis* (gi|22164316). Three related sequences were also described in *I. pacificus*. We here report seven additional sequences from *I. scapularis* (Figure 10), indicating that the 18.7-kDa proteins also belong to a multigene family consisting of at least two polymorphic genes. The *I. pacificus* sequences cluster together into one clade and have in common an insertion (marked by the box in Figure 10A) containing two additional cysteines. Only a framework of mostly conserved cysteines, glycines and prolines and a few additional AA identities and similarities group this diverse family together. The function of this protein family is unknown.

Group 6: similar to I. scapularis *5.3-kDa secreted peptide.* This multigene peptide family (Figure 11) has six conserved cysteine residues and codes for peptides with mature molecular weight near 5 kDa. This family is also found in *I. pacificus* and has no known function; however, members of this family were significantly increased in transcription in the salivary glands of nymphs infected with *B. burgdorferi* (Table 4A), suggesting it may be related to antimicrobial defense mechanisms. The paralytic salivary toxin of the Australian tick *Ixodes holocyclus* produces a poor match to one member of the 5.3-kDa peptide family, to which it may be distantly related. No other match of holocyclotoxin was found to our database, suggesting the Australian tick toxin is not found in *I. scapularis*. The Edman degradation result, E-[PV]-E-P-G-X-A-Y-Q-V-K-A-G-R from a SDS-PAGE protein band separating saliva (Valenzuela et

al., 2002) identified gi|22164302 with a match EPdPGqpwQVKAGR at position 22 of the native peptide and expected cleavage site of the signal peptide (lower case designates mismatch; results in Supplemental Table 2), indicating this or a related peptide is found in saliva of *I. scapularis*.

Group 7: 9- and 7-kDa families of peptides. This family includes subfamilies of 7- and 9-kDa (Figure 12) that are distantly related. They contain a conserved framework of four or eight cysteines, depending on the final size (Figure 12A). Homologs are also found in *I. pacificus*, but not on other ticks. The family appears to have existed before separation of *I. scapularis* and *I. pacificus* (Figure 12B). Some members of this family of unknown function resemble scorpion toxins, mainly due to a cysteine framework. Five members of this family were identified in saliva by Edman degradation of protein bands separated by SDS-PAGE, indicating that members of this peptide family are expressed.

Group 8: metalloproteases. Metalloproteases with fibrin(ogen)lytic activity in *I. scapularis* saliva, where they may function as an antihemostatic and anti-angiogenic factor, have been described previously (Francischetti et al., 2003; Packila and Guilfoile, 2002; Valenzuela et al., 2002). For further description of this protein family, consult (Francischetti et al., 2003). Members of this enzyme family are abundantly expressed, as evidenced by strong Edman degradation signal of saliva separated by SDS-PAGE (Valenzuela et al., 2002). The phylogram in Figure 13 indicates the multi gene nature of this family.

Group 9: GPIIb/IIIa antagonists. Arg-Gly-Asp (RGD) containing polypeptides flanked by cysteines forming a disulfide bridge produce a hairpin loop that is known to interact with integrins, proteins involved in cellular attachment, including platelet aggregation (Niewiraowski et al., 1994). Similar peptides containing Lys-Gly-Asp (KGD) instead of RGD are much more specific for the platelet integrin, the glycoprotein (GP) IIb-IIIa (Scarborough et al., 1993). Seven such proteins are represented in *I. scapularis* transcriptomes, five of which are similar to a

peptide also found in *I. pacificus* (Figure 14A) and have been described in detail (Francischetti et al., 2005). Two other unrelated sequences have either a RGD or a KGD domain and an additional XGD motif flanked by cysteines (Figure 14B). The design of this second hairpin XGD loop suggests a double interaction of this novel family of disintegrins to a classical integrin site and to another yet-unidentified target. Another very similar putative peptide was found that lacks the RGD or KGD motif, having instead a Ser-Gly-Asp motif. Notably, SGD motifs have been shown to interact with extracellular matrix glycosaminoglycans in perlecan (Dolan et al., 1997). These proteins most certainly inhibit platelet aggregation and may also affect other cellular interactions.

Group 10: the ixostatin family, or short-coding cysteine-rich-peptides. This family codes for cysteine-rich peptides of mature molecular weight varying from 9–11 kDa that are similar to the cysteine-rich domains of ADAMTS (a disintegrin and metalloprotease with thrombospondin motifs) proteins, which are enzymes that cleave cartilage and inhibit angiogenesis, among other functions (Porter et al., 2005). Some of the members have the Pfam signature of ACD, for ADAMS cysteine-rich domain. This family includes one member misnamed as thrombospondin. Thirty-eight protein sequences representing this family in *I. scapularis* are shown in Supplemental Table 2 and Figure 15 (30 of which are new), together with similar proteins from *I. pacificus*. The phylogram suggests that the majority of the members of this group might have evolved from a common ancestor and that multiple genes are represented in this protein family. The function of this protein family is unknown, but it might be involved with angiogenesis/repair inhibition, as indicated elsewhere (Francischetti et al., 2005).

Group 11: lipocalins. Lipocalins are ubiquitous proteins of ~20 kDa that specialize in binding and/or transporting small ligands (Flower et al., 2000). In the blood-sucking *Rhodnius prolixus*, there are several different salivary lipocalins specialized in transporting nitric oxide,

and binding histamine, serotonin, norepinephrine, and adenine nucleotides (Andersen et al., 2005). In hard ticks, some members of this family were shown to bind histamine and serotonin (Sangamnatdej et al., 2002), while others are toxic in soft ticks (Mans et al., 2003). The family diversification in ticks is evident from the complex phylogram (Figure 16) containing 52 protein sequences from *I. scapularis* (39 novel) (Figure 16A), and an additional 34 sequences from *I. pacificus, R. appendiculatus*, *B. microplus, Hyalomma asiaticum, Dermacentor reticulatus*, and the soft ticks *Argas reflexus* and *Ornithodorus savignyi*, which may derive from three main common ancestors (Figure 16B). For a discussion of the evolution of this gene family in ticks, see references (Mans and Neitz, 2004a; Mans and Neitz, 2004b; O'Neill et al., 1995).

Group 12: neuropeptide-like polypeptides with GGY repeats and other antimicrobial peptides. The GGY family contains basic peptides of mature molecular weight ranging from 4.7-13 kDa and having sequence similarity to a *Caenorhabditis elegans* peptide family previously identified as neuropeptides (Li et al., 1999; Nathoo et al., 2001) but more recently identified as having potent antimicrobial activity (Couillault et al., 2004). Members of this family exist also in other tick species (Francischetti et al., 2005). Two other unrelated families are also included in Group 12: the previously reported salivary defensin (gi|56159961), a widespread family of antimicrobial peptides (Torres and Kuchel, 2004), that also contain a single GGY motif in their carboxyterminus; and two novel sequences related to the previously described bacteriostatic peptide microplusin from the tick *B. microplus* containing a AHHE carboxy signature (Fogaca et al., 2004). Antimicrobial peptides secreted in tick saliva may prevent infection and disruption of feeding at the immunosuppressed feeding site. It is also interesting that defensin polypeptides are know to affect cytokine production (Chaly et al., 2000; Sakamoto et al., 2005) and could act as immunomodulators at the tick feeding site.

Group 13: oxidant metabolism. This group includes glutathione peroxidases and selenoproteins that may be implicated in preventing inflammatory oxidant reactions at the tick feeding site. A salivary glutathione peroxidase was previously reported from *I. scapularis* coded by gi|15428288 (Das et al., 2001). This peroxidase does not have a signal peptide indicative of secretion but was identified as an immunodominant antigen in *I. scapularis* saliva (Das et al., 2001). We report here another glutathione peroxidase unrelated to gi:15428288 but having a signal peptide indicative of secretion and similar to human and rat plasma glutathione peroxidases. This type of enzyme is the main plasma antioxidant in mammals and has considerable anti-hemostatic activity by increasing the half life of NO (Jin et al., 2005). Three other putative protein sequences are similar to thioredoxin peroxidases, which are proteins of the same family as glutathione peroxidase. These also have signal peptide indicative of secretion, although two of the thioredoxin peroxidases are similar to mitochondrial enzymes, and that may be their final destination instead of the saliva.

Selenoproteins include in their composition the rare AA selenocysteine (Sec). For the incorporation of selenocysteine into protein, the UGA codon is transformed from one that signals translation termination to one specific for selenocysteine (Copeland, 2003), when selenium is available. The family of selenoproteins includes a short group of ~15 kDa that have no known function but are thought to be involved in oxidant metabolism (Beckett and Arthur, 2005). Significantly, *I. scapularis* salivary transcriptome reveals three distinct short selenoproteins, two of which have a signal peptide indicative of secretion (Supplemental Table 2). Their role in feeding remains to be investigated.

Group 14: anticomplement proteins (Isac). Saliva of adult female *I. scapularis* inhibits the alternative pathway of complement activation (Ribeiro, 1987). A salivary protein named Isac (gi|8896135) was identified and the expressed recombinant protein shown to act by interfering

with factor B binding to C3 convertase (Valenzuela et al., 2000). *I. ricinus* salivary gland homogenate also interfere with the alternative pathway of complement activation, preventing C3 deposition (Lawrie et al., 2005). Supplemental Table 2 reports seven sequences of this family from *I. scapularis*, four of which are novel. The previously reported protein salp9 (gi|15428346) does not have a signal peptide and most likely reflects sequencing of a truncated clone. All members of the *I. scapularis* Isac family have the carboxyterminus rich in Ser and Thr AA to which 8–14 sites indicative of O-galactosylation usually map. This glycosylation pattern may explain the anomalous chromatographic behavior of the originally purified anticomplement protein (Valenzuela et al., 2000). Homologs of Isac occur in *I. pacificus* and *I. ricinus*. The alignments and phylogram suggests that these sequences derive from a multigene family (Figure 17). Edman degradation of protein gel bands in SDS-PAGE experiments identified sequences belonging to this family starting at the predicted cleavage site of the signal peptide.

Group 15: the WC-10 peptide family. This redundant family expresses putative secreted peptides with an average mass of 9–11 kDa and having in common a conserved WC dipeptide motif in the carboxyterminus or similarity to peptides with a WC motif (Figure 18). This family appears to be unique to *I. scapularis*. Its function is unknown.

Group 16: the LPTS peptide family. This peptide family has a mildly acidic pI, the mature product ranges from 12–16 kDa, and most members have a [LVI]-P-[TS]-C motif on its carboxyterminus region. Most have more than four predicted O-galactosylation sites scattered at the middle and carboxyterminus of the molecule. Related products are also found in *I. pacificus*, but not on other ticks, so far. This family appears also to be multigenic (Figure 19).

Group 17: other putative antiprotease polypeptides. This group includes serpins, cystatins, and trypsin inhibitor-like (TIL) domain-containing peptides. Serpins are a widespread family of serine protease inhibitors, of which two sequences (gi|22164288 and gi|56159953)

have been reported before following a sialotranscriptome study of *I. scapularis* (Valenzuela et al., 2002). These inhibitors may target clotting or complement enzymes. A salivary cystatin, a member of a protein family of inhibitors of cysteine proteases, has also been previously described (gi|22164282). Table 8 reports an additional cystatin that is 75% identical to the previously reported cystatin. These inhibitors may target antigen-presenting cells, cathepsins, or extracellular cathepsins involved in tissue repair (Abrahamson et al., 2003; Nakagawa and Rudensky, 1999; Vray et al., 2002). TIL domain peptides have not been described before in the *I. scapularis* sialotranscriptome. Table 8 presents two protein sequences, possibly alleles, that have similarities to peptides annotated as containing the TIL domain. These putative proteins may target serine proteases.

Group 18: mucins. This heterogeneous group includes five novel protein sequences having in common 14 or more O-galactosylation sites and two possibly allelic sequences containing four O-galactosylation sites and chitin-binding domains. The chitin-binding domains may direct these proteins to line the chitinous mouthparts of the tick. These putative glycoproteins may function in maintenance of the tick mouthparts.

Group 19: the IS6 family. This family contains secreted peptides of mature molecular weight ranging from 9–12 kDa, of unknown function. The cysteine framework matches peptides annotated as neurotoxins, disintegrins, or the hormone prokineticin. No similar sequences are found in any other tick species, although *I. pacificus* has one sequence that has weak similarity to the group. These peptides may function as vasodilators, neurotoxins, or antimicrobials.

Group 20: the 12-kDa family. This family of three known members contains similar protein in *I. pacificus* and codes for mature acidic peptides ranging from 12 to 13 kDa. The function is unknown.

Group 21: the 26-kDa family. This family is unique to *I. scapularis* and contains two previously reported acidic proteins of 23 and 26 kDa sharing less than 30% identity. The function is unknown.

Group 22: the 30-kDa family. This group reports two novel putative secreted sequences, possibly allelic, sharing weak sequence similarity with bacterial toxins annotated as cytotoxins. Similar sequences occur in *I. pacificus.*

Group 23: toxin like. Two novel sequences are reported in this group, coding for predicted mature peptides of 8–9 kDa. They share substantial similarity on their signal peptide, but the mature peptides are quite divergent, with two and six cysteines, one of which has weak similarity to peptides annotated as toxins. They are also weakly related to the IS6 family(Group 19) above, which has similar signal peptide sequences.

Group 24: the SRAEL family. Two novel sequences compose this family of putative secreted peptides of 16–22 kDa and having in their carboxyterminus the SRAEL sequence. These sequences are unique to *I. scapularis* and have no known function.

Group 25: other enzymes. In addition to the metalloproteases of Group 8, several other enzymes are found to be possibly secreted in *I. scapularis* saliva, including one previously described serine carboxypeptidase (gi|22164290) that may actually be a lysosomal enzyme, and five enzymes, as follows:

- Two trypsin-like enzymes, one novel that has similarity to T-plasminogen activator of *Sus scrofa,* and is similar to a previously reported salivary protein of fed ticks in *I. scapularis* (gi|55736035). These serine proteases may be involved in clotting or inflammation, such as plasminogen activation or protein C activation.
- Two dipeptidyl peptidases (both are truncated sequences), similar to angiotensin-converting enzyme or endothelin converting enzyme, that could be

responsible for the reported salivary kininase activity of *I. scapularis* (Ribeiro and Mather, 1998). This enzyme is part of a multi-family as indicated by many EST coding for similar enzyme sequences (Supplemental Table 1).

- A phospholipase A_2, also truncated, indicating that possibly *I. scapularis* may contain this type of activity, as was reported for Amblyomma (Zhu et al., 1998).
- A secreted RNAse belonging to a ubiquitous family of small RNAses (Rampias et al., 2003).

Fragments of 5/ nucleotidases are also found in the EST collection, comprising seven contigs; one or more of these may function as a salivary apyrase, as is the case in mosquitoes (Champagne et al., 1995). These contigs are annotated as 5/-nucleotidase–apyrase in the comments column of Supplemental Table 1. Full-length sequences for these EST were not obtained in this work.

Group 26: other polypeptides of unknown function. An additional 42 peptides and proteins are listed in Supplemental Table 2. These cannot be grouped within related families and most have no known function. Of note are:

- a putative secreted protein containing a fibrinogen-binding motif similar to ficolin and angiopoietin and to a *I. ricinus* protein named ixoderin B (gi|58422507). This protein may be involved in matrix attachment processes and angiogenesis inhibition. Alternatively, it may antagonize the effect of host ficolin, a protein involved in the activation of the lectin-complement pathway (Fujita et al., 2004).
- Another novel putative secreted protein belongs to the antigen 5 family, which is found in wasp venom and in many hematophagous insect transcriptomes (Ribeiro and Francischetti, 2003).

The function for the vast majority of members of this ubiquitous protein family is not known (King and Spangfort, 2000); however, in a venomous lizard and snake venom, the proteins named helothermine (Mochca-Morales et al., 1990) and ablomin (Yamazaki et al., 2002) inhibit ion channel activity and could potentially act as vasodilators or neurotoxins. A totally different function was found in a member of this family that has a specific proteolytic activity in processing propeptide precursors of Conus toxins (Milne et al., 2003). The role(s) of this family of proteins in the saliva of blood-sucking arthropods remains to be identified.

It is possible that some of the proteins in this group may perform a housekeeping function at intracellular compartments instead of being secreted, participating in endoplasmic reticulum or Golgi functions.

Group 27: H proteins. We here report 167 novel sequences from *I. scapularis* that we believe play a housekeeping role, adding to 14 previously reported sequences. These are grouped together in the "Housekeeping" worksheet of Supplemental Table 2. Many ribosomal proteins and other products associated with protein modification and export are reported. Of interest to the function of the salivary glands are sequences associated with posttranslation modification, including the discovery of cDNA sequences coding for proline hydroxylase, aryl sulfotransferase, and an enzyme subunit involved in vitamin K-dependent formation of γ-carboxyglutamate, which will be described in detail.

- Proline hydroxylase enzymes such as procollagen-proline, 2-oxoglutarate-4-dioxygenase catalyze the posttranslation modification found in collagen. The clone IS-6-12-J-cluster-143 codes for a protein similar to members of the α subunit of this enzyme, indicating that hydroxylation of proline residues in tick saliva is possible.
- Aryl sulfotransferases posttranslationally modify tyrosine residues in eukaryotic cells. Sulfotyrosine modification occurs in several hormones (Cantor et al., 1986; Eng et al.,

1986; Kohli et al., 1988), clotting factors such as Factor VIII (Pittman et al., 1992), and clotting inhibitors such as hirudin (Niehrs et al., 1990), where the activity is significantly enhanced by the modification. Similarly, venomous peptides from the mollusk Conus are also sulfated (Loughnan et al., 1998). Two members of the arylsulfatase family were found in *I. scapularis,* those coding for ISUFJ-cluster-392 and ISUFJ-cluster-165, which are distantly related, suggesting an ancient gene duplication for this gene or fast evolution after duplication. The finding of transcripts coding for this enzyme activity in tick salivary glands suggests that sulfation of tyrosine residues may occur in tick salivary polypeptides.

- The protein sequence predicted by ISUFL301 is similar to many proteins annotated as subunit 1 of the vitamin K epoxide reductase complex. This subunit is fundamental to posttranslational carboxylation of glutamate to form the γ-carboxyl glutamic acid residues found in blood-clotting serine proteases and other proteins active in hemostasis and inflammation (Furie et al., 1999; Zhang and Ginsburg, 2004). This modification is essential for the proper assembly of procoagulation complexes, such as the Xase and prothrombinase complexes (Burnier et al., 1981). Of interest, venomous peptides from the mollusk Conus also have glutamic acid residues (Bandyopadhyay et al., 1998; McIntosh et al., 1984). It is thus possible that some tick salivary proteins may contain glutamic acid residues that may interfere with assembly of procoagulation protein complexes.

To the extent that salivary polypeptides are posttranslationally modified by these enzymes and that host antibodies reach the salivary glands, immunologic targeting of these enzymes could affect production of whole families of peptides and proteins with a single vaccine target. Supplemental Table 2 shows the number of potential sites where sulfotyrosine could

occur, as well as the number of potential proline hydroxylations, and glutamate carboxylations, based on known eukaryotic sequence specificity for the enzymes (de Jong et al., 1991; Nicholas et al., 1999). These posttranslation modifications, as well as glycosylations (see gi|56159969), should be taken into consideration if recombinant or synthetic proteins are attempted for vaccine or pharmacologic studies. Potential hydroxylation of the proline-rich peptides of Group 4, as occurs with collagen, has been mentioned above. The motif known to trigger vitamin K-dependent glutamate modification in mammals, E-X(3)-E-X-C (Price et al., 1987) occurs in the majority of Group 1 (basic tail) peptides, from which salp14 has been demonstrated to be anticlotting (Narasimhan et al., 2002), and also occur in Group 2 peptides (Supplemental Table 2).

Concluding Remarks

With the completion of this transcriptome analysis, it becomes evident that gene expression in the salivary glands of *I. scapularis* is very complex. For comparison, a recent analysis of *An. gambiae* sialotranscriptome based on ~3000 EST produced nearly 70 proteins that are putatively secreted in adult mosquito salivary glands (Ribeiro et al., manuscript submitted), while in this manuscript, we cataloged 470 putative secreted salivary proteins, which are grouped in more than 25 protein families. While mosquitoes feed for only a few minutes, ticks feed for several days or weeks and thus have to cope with the host late-inflammatory (wound-healing) and cellular immune reactions. The protein salivary repertoire of ticks must then be more complex than that of fast-feeding insects.

As indicated previously (Mans and Neitz, 2004a), many tick salivary proteins belong to families that appeared to have evolved by gene duplication events. Gene duplication may account for 8% to 25% of all genes in different genomes (Sankoff, 2001). Duplicated genes

usually accumulate enough mutations to become inactive pseudogenes, but they may become fixed if they are of adaptive value. Duplicated genes may persist if they advantageously increase the dosage of a transcript in a particular tissue, if they evolve to advantageously express the gene in a tissue-selective way, or if they advantageously diverge from the original gene creating a new, but related, function (Mazet and Shimeld, 2002). All these mechanisms may have occurred in ticks to promote fixation of duplicated genes. It is also possible that some duplicated genes have changed from the original to obtain antigenic variation while maintaining the same function, as occurs with the alleles of maxadilan, the vasodilator of the sand fly *Lutzomyia longipalpis* (Milleron et al., 2004). For example, if we assume all the basic tail members are anticlotting in nature, the many duplicated genes of this family may have become fixed by having increased the expression of the polypeptides in saliva (dosage effect), by having evolved to target different clotting enzymes or to target the same enzyme in different vertebrate hosts (divergence of function effect) or to achieve antigenic variation without change in host enzyme target (antigenic variation effect), or to have their expression selected at different instars or at different times of feeding, as is the case of salp9, indicated above to be expressed mostly in nymphs. Many of these evolutionary scenarios imply changes also in the gene regulatory sequences. Antigenic variation appears to be an important selective pressure in genes coding for salivary proteins, as indicated by the polymorphism that appears in high frequency at many loci, suggesting a frequency-dependent selection scenario driven by host immunity against salivary protein epitopes. Switching gene expression of antigenically different members of the same family may be a tick strategy to achieve a successful blood meal. This scenario is consistent with the life history of *I. scapularis*. Newborn larvae feed mostly on *Peromyscus leucopus* mice within a 2- to 3-week period, in late August, when the majority of the summer-born rodent population have not experienced tick bites. By the time the mice evolve any immune response,

most of the larvae had already fed. Larvae molt to nymphs, which feed the following year, in late May, on the same mouse population that survived the winter (Spielman et al., 1985). It would be advantageous to nymphs to switch salivary cocktails. Adults, on the other hand, feed in late fall and early spring, mostly on middle and large size mammals, mostly deer. This timing strategy, however, may be a relic of early evolution. Adaptation of *I. scapularis* ticks to feed on mice leads to a virtually complete immunoevasion of host responses. *P. leucopus* immunized by repeated tick bites cannot reject *I. scapularis* (Davidar et al., 1989; Ribeiro, 1989), although non-natural hosts such as the guinea pig do so efficiently (Trager, 1939). A differential salivary cocktail in larval ticks remains to be determined, and although we await development of suitable techniques to deal with such minute organisms, gene duplication events may have nonetheless facilitated the adaptation of ticks to blood feeding.

Gene duplication may occur by several mechanisms, including genome duplication (leading to polyploidy), tandem duplication (due to illegitimate recombinational events), and transposition (Sankoff, 2001). The role of genome duplication in evolution has been proposed by Ohno (Ohno, 1970; Ohno, 1999) to explain the evolution of vertebrates. Based on cytogenetic data, two rounds of genome duplication were proposed to have occurred as nonvertebrate chordates evolved to primitive fishes, such as the lamprey, and then to tetraploid organisms, comprising all gnathostome (gnaw-containing) vertebrates (Ohno, 1970). Ohno's hypothesis was substantiated by the finding that many regulatory genes, such as the homeobox gene cluster (controlling embryonic segment development) are increased in number in modern vertebrates when compared to more primitive ancestral organisms; however, the hypothesis has been challenged by the proposition that many duplications in such key developmental genes have occurred after the tetraploidy occurrence by a mechanism of tandem gene duplication (Holland, 1999). On the other hand, because the vertebrate tetraploidy event occurred over 400 million

years ago, chromosomal rearrangements, further gene tandem duplications, and gene silencing could have obscured the gene arrangements following such a major, catastrophic, evolutionary event. Ticks are specialized mites, organisms known usually to have very few chromosome numbers (Oliver, 1977; Oliver and Nelson, 1967). Within the Parasitiformes, the lineage containing ticks, the Mesostigmata are the sister group, and these contain typically 4–5 chromosome pairs (Hansell et al., 1964; Wysoki and Swirski, 1968). Hard ticks (Ixodidae) appear to have evolved from bird-feeding soft ticks similar to *Argas* (Black and Piesman, 1994). Most have 13 pairs of chromosomes, while *I. scapularis* has 14 chromosome pairs (Chen et al., 1994). It thus appears that early in tick evolution genome duplications occurred once, perhaps twice, from the ancestral mite. Adaptation to blood feeding must have created a natural selection reward for the evolution of a salivary cocktail able to disarm host hemostasis, inflammation, and immunity; as such, genome duplications must have created an "instantaneous" increase in the substrate for evolution and diversification of these compounds, in addition to a possible immediate effect of increased salivary gland dosage. Unraveling of the *I. scapularis* genome, which is an ongoing project (Hill and Wikel, 2005), will allow unique insight into the role of genome duplications in evolution. Because these duplication events must have occurred much after the period of gnathostome genome duplication and, thus, their tracks should be less erased by time, and because Ixodes has such a large number of duplicated genes associated with salivary function, their adaptive role on feeding can be estimated and manipulated (for example, with RNAi technology).

Other blood-sucking arthropods also show evidence of gene families coding for salivary proteins. This is the case for the D7 (with eight genes in tandem repeats) (Arca et al., 2002) and the g1 (four genes in tandem repeats) families of salivary proteins in *An. gambiae* (Ribeiro et al., manuscript submitted). Similarly, *Rhodnius* salivary lipocalins have undergone a

still larger expansion than that seen for mosquito salivary proteins (Andersen et al., 2005; Ribeiro et al., 2004), and compares with the expansion seen in some *Ixodes* protein families; however, ticks have many more expanded protein families than are known in *Rhodnius*. Of interest, mosquitoes have only 3 chromosome pairs, while triatomines have 11 pairs (Panzera et al., 1995). Non-blood-sucking Heteroptera, however, have similar number of chromosomes, indicating that genome duplication in Triatomines was not associated with the evolution to blood feeding; however, a larger genome in *Rhodnius* compared with Anopheles might have allowed a larger substrate for evolution of tandem repeated genes associated with blood feeding. The genome of *R. prolixus* has been also recently targeted for sequencing and should give additional insights into the evolution of blood feeding by arthropods.

In addition to genome duplications (manifested by the presence of related families of genes in different chromosomes or related gene order in different chromosomes), tandem duplications and possible retrotransposition events (evidenced by the several transcripts indicative of retrotransposase) might also have contributed to evolution of Ixodes to blood feeding. Recently, the genome size of several ticks, including *I. scapularis* has been determined (Ullmann et al., 2005) and revealed a large proportion of mid-repetitive DNA, supporting the existence of an unusual amount of tandem duplications in tick genomes. With the completion of the *I. scapularis* genome, it will be interesting to compare the tandem-repeated genes of the same family but located in different chromosomes. Gene conversion events may be a barring force in the evolution of tandem repeats; thus, repeats in different chromosomes may diverge more quickly than sequences in the same chromosome. Did the repeats appear before or after genome duplication, or both? How conserved are they within and between chromosomes? Single exon genes might indicate retrotransposition events. Could it be that retrotransposition of signal peptide sequences containing salivary promoter regions “land” in intronic regions of other genes,

creating new salivary proteins? Genome duplication events produce many spare targets for such retrotransposon landings, which would otherwise disrupt essential genes. These and many other questions wait deciphering of the *I. scapularis* genome.

Last but not least is the perplexing number of genes that appear to code for secreted polypeptides but for which we have no insight into their function. The unknown can be divided into problems, where we have the basic knowledge to propose a logical solution, and mysteries, where we have no basis to formulate a solution. While the Kunitz-containing peptides are in the first category (they must function, as a rule, by inhibiting serine proteases), most other gene families described above are mysterious. These must play a role in the host physiology and biochemistry, but we do not have enough information to predict and test their effects. The pharmacology and biochemistry textbooks of today are the result of 200 years of human enquiry, while blood-sucking arthropods have in their saliva the result of at least 100 million years of evolution. Expression and high-throughput bioassay screening of these molecules must lead us to a better understanding of tick biology and of human physiology.

Acknowledgments

This study was supported in part by award number DAMD17-03-1-005 from The United States Army Medical Research and Materiel Command to S.K.W. We are grateful to Ms. Brenda Marshall for editorial assistance.

References

Abrahamson M, Alvarez-Fernandez M, Nathanson CM, 2003. Cystatins. Biochem Soc Symp:179-99.

Altschul SF, Gish W, 1996. Local alignment statistics. Methods Enzymol 266:460-80.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ, 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-402.

Andersen JF, Gudderra NP, Francischetti IM, Ribeiro JM, 2005. The role of salivary lipocalins in blood feeding by *Rhodnius prolixus*. Arch. Insect Biochem. Physiol. 58:97-105.

Andersen JF, Gudderra NP, Francischetti IM, Valenzuela JG, Ribeiro JMC, 2004. Recognition of anionic phospholipid membranes by an antihemostatic protein from a blood-feeding insect. Biochemistry 43:6987-94.

Anderson JF, 2002. The natural history of ticks. Med Clin North Am 86:205-18.

Anguita J, Ramamoorthi N, Hovius JW, Das S, Thomas V, Persinski R, Conze D, Askenase PW, Rincon M, Kantor FS, Fikrig E, 2002. Salp15, an *Ixodes scapularis* salivary protein, inhibits CD4(+) T cell activation. Immunity 16:849-59.

Arca B, Lombardo F, Lanfrancotti A, Spanos L, Veneri M, Louis C, Coluzzi M, 2002. A cluster of four D7-related genes is expressed in the salivary glands of the African malaria vector *Anopheles gambiae*. Insect Mol. Biol. 11:47-55.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G, 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25:25-9.

Bandyopadhyay PK, Colledge CJ, Walker CS, Zhou LM, Hillyard DR, Olivera BM, 1998. Conantokin-G precursor and its role in gamma-carboxylation by a vitamin K-dependent carboxylase from a Conus snail. J Biol. Chem 273:5447-50.

Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL, 2000. The Pfam protein families database. Nucleic Acids Res. 28:263-6.

Beckett GJ, Arthur JR, 2005. Selenium and endocrine systems. J Endocrinol 184:455-65.

Black WCt, Piesman J, 1994. Phylogeny of hard- and soft-tick taxa (Acari: Ixodida) based on mitochondrial 16S rDNA sequences. Proc. Natl. Acad. Sci. U S A. 91:10034-8.

Black WCt, Roehrdanz RL, 1998. Mitochondrial gene order is not conserved in arthropods: prostriate and metastriate tick mitochondrial genomes. Mol Biol. Evol 15:1772-85.

Broze GJ, Girard TJ, Novotny WF, 1990. Regulation of coagulation by a multivalent Kunitz-type inhibitor. Biochemistry 29:7539-7546.

Burnier JP, Borowski M, Furie BC, Furie B, 1981. Gamma-carboxyglutamic acid. Mol Cell Biochem 39:191-207.

Cantor P, Andersen BN, Rehfeld JF, 1986. Complete tyrosine O-sulfation of gastrin in adult and neonatal cat pancreas. FEBS Lett 195:272-4.

Chaly YV, Paleolog EM, Kolesnikova TS, Tikhonov, II, Petratchenko EV, Voitenok NN, 2000. Neutrophil alpha-defensin human neutrophil peptide modulates cytokine production in human monocytes and adhesion molecule expression in endothelial cells. Eur Cytokine Netw 11:257-66.

Champagne DE, Smartt CT, Ribeiro JM, James AA, 1995. The salivary gland-specific apyrase of the mosquito *Aedes aegypti* is a member of the 5'-nucleotidase family. Proc. Natl. Acad. Sci. U S A 92:694-8.

Chen C, Munderloh UG, Kurtti TJ, 1994. Cytogenetic characteristics of cell lines from Ixodes scapularis (Acari: Ixodidae). J Med Entomol 31:425-34.

Copeland PR, 2003. Regulation of gene expression by stop codon recoding: selenocysteine. Gene 312:17-25.

Couillault C, Pujol N, Reboul J, Sabatier L, Guichou JF, Kohara Y, Ewbank JJ, 2004. TLR-independent control of innate immunity in Caenorhabditis elegans by the TIR domain adaptor protein TIR-1, an ortholog of human SARM. Nat Immunol 5:488-94.

Das S, Banerjee G, DePonte K, Marcantonio N, Kantor FS, Fikrig E, 2001. Salp25D, an *Ixodes scapularis* Antioxidant, Is 1 of 14 Immunodominant Antigens in Engorged Tick Salivary Glands. J. Infect. Dis. 184:1056-64.

Davidar P, Wilson M, Ribeiro JM, 1989. Differential distribution of immature Ixodes dammini (Acari: Ixodidae) on rodent hosts. J Parasitol 75:898-904.

de Jong L, van der Kraan I, de Waal A, 1991. The kinetics of the hydroxylation of procollagen by prolyl 4-hydroxylase. Proposal for a processive mechanism of binding of the dimeric hydroxylating enzyme in relation to the high kcat/Km ratio and a conformational requirement for hydroxylation of -X-Pro-Gly- sequences. Biochim Biophys Acta 1079:103-11.

Dolan M, Horchar T, Rigatti B, Hassell JR, 1997. Identification of sites in domain I of perlecan that regulate heparan sulfate synthesis. J Biol. Chem 272:4316-22.

Eng J, Gubler U, Raufman JP, Chang M, Hulmes JD, Pan YC, Yalow RS, 1986. Cholecystokinin-associated COOH-terminal peptides are fully sulfated in pig brain. Proc. Natl. Acad. Sci. U S A 83:2832-5.

Flower DR, North AC, Sansom CE, 2000. The lipocalin protein family: structural and sequence overview. Biochim. Biophys. Acta 1482:9-24.

Fogaca AC, Lorenzini DM, Kaku LM, Esteves E, Bulet P, Daffre S, 2004. Cysteine-rich antimicrobial peptides of the cattle tick Boophilus microplus: isolation, structural characterization and tissue expression profile. Dev Comp Immunol 28:191-200.

Francischetti IM, Mather TN, Ribeiro JM, 2003. Cloning of a salivary gland metalloprotease and characterization of gelatinase and fibrin(ogen)lytic activities in the saliva of the Lyme disease tick vector *Ixodes scapularis*. Biochem Biophys Res Commun 305:869-75.

Francischetti IM, Mather TN, Ribeiro JM, 2004. Penthalaris, a novel recombinant five-Kunitz tissue factor pathway inhibitor (TFPI) from the salivary gland of the tick vector of Lyme disease, Ixodes scapularis. Thromb Haemost 91:886-98.

Francischetti IM, Valenzuela JG, Andersen JF, Mather TN, Ribeiro JM, 2002a. Ixolaris, a novel recombinant tissue factor pathway inhibitor (TFPI) from the salivary gland of the tick, *Ixodes scapularis*: identification of factor X and factor Xa as scaffolds for the inhibition of factor VIIa/tissue factor complex. Blood 99:3602-12.

Francischetti IM, Valenzuela JG, Pham VM, Garfield MK, Ribeiro JM, 2002b. Toward a catalog for the transcripts and proteins (sialome) from the salivary gland of the malaria vector *Anopheles gambiae*. J. Exp. Biol. 205:2429-51.

Francischetti IMB, Pham VM, Mans BJ, Andersen JF, Mather TN, Lane RS, Ribeiro JMC, 2005. The transcriptome of the salivary glands of the female western black legged tick *Ixodes pacificus* (Acari: Ixodidae). Gene In press.

Fujita T, Matsushita M, Endo Y, 2004. The lectin-complement pathway--its role in innate immunity and evolution. Immunol Rev 198:185-202.

Furie B, Bouchard BA, Furie BC, 1999. Vitamin K-dependent biosynthesis of gamma-carboxyglutamic acid. Blood 93:1798-808.

Gillespie RD, Dolan MC, Piesman J, Titus RG, 2001. Identification of an IL-2 binding protein in the saliva of the Lyme disease vector tick, *Ixodes scapularis*. J. Immunol. 166:4319-26.

Gillespie RD, Mbow ML, Titus RG, 2000. The immunomodulatory factors of bloodfeeding arthropod saliva. Parasite Immunol 22:319-31.

Gregson JD, 1967. Observations on the movement of fluids in the vicinity of the mouthparts of naturally feeding Dermacentor andersoni Stiles. Parasitol. 57:1-8.

Hansell RI, Mollison MM, Putman WL, 1964. A Cytological Demonstration of Arrhenotoky in Three Mites of the Family Phytoseiidae. Chromosoma 15:562-7.

Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, Brunak S, 1998. NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. Glycoconj J 15:115-30.

Hill CA, Wikel SK, 2005. The Ixodes scapularis Genome Project: an opportunity for advancing tick research. Trends Parasitol 21:151-3.

Holland PW, 1999. Gene duplication: past, present and future. Semin Cell Dev Biol 10:541-7.

Huang X, Madan A, 1999. CAP3: A DNA sequence assembly program. Genome Res 9:868-77.

Jin RC, Voetsch B, Loscalzo J, 2005. Endogenous mechanisms of inhibition of platelet function. Microcirculation 12:247-58.

Kaufman WR, 1989. Tick-host interaction: a synthesis of current concepts. Parasitol Today 5:47-56.

King TP, Spangfort MD, 2000. Structure and biology of stinging insect venom allergens. Int Arch Allergy Immunol 123:99-106.

Kivirikko KI, Kishida Y, Sakakibara S, Prockop DJ, 1972. Hydroxylation of (X-Pro-Gly)n by protocollagen proline hydroxylase. Effect of chain length, helical conformation and amino acid sequence in the substrate. Biochim Biophys Acta 271:347-56.

Knight CG, Morton LF, Onley DJ, Peachey AR, Ichinohe T, Okuma M, Farndale RW, Barnes MJ, 1999. Collagen-platelet interaction: Gly-Pro-Hyp is uniquely specific for platelet Gp VI and mediates platelet activation by collagen. Cardiovasc Res 41:450-7.

Kohli R, Chadha N, Muralidhar K, 1988. Presence of tyrosine-O-sulfate in sheep pituitary prolactin. FEBS Lett 242:139-43.

Kumar S, Tamura K, Nei M, 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 5:150-63.

Lawrie CH, Sim RB, Nuttall PA, 2005. Investigation of the mechanisms of anti-complement activity in Ixodes ricinus ticks. Mol Immunol 42:31-8.

Lewis S, Ashburner M, Reese MG, 2000. Annotating eukaryote genomes. Curr. Opin. Struct. Biol. 10:349-54.

Li C, Nelson LS, Kim K, Nathoo A, Hart AC, 1999. Neuropeptide gene families in the nematode Caenorhabditis elegans. Ann N Y Acad Sci 897:239-52.

Loughnan M, Bond T, Atkins A, Cuevas J, Adams DJ, Broxton NM, Livett BG, Down JG, Jones A, Alewood PF, Lewis RJ, 1998. alpha-conotoxin EpI, a novel sulfated peptide from Conus episcopatus that selectively targets neuronal nicotinic acetylcholine receptors. J Biol. Chem 273:15667-74.

Mans BJ, Louw AI, Neitz AW, 2003. The major tick salivary gland proteins and toxins from the soft tick, Ornithodoros savignyi, are part of the tick Lipocalin family: implications for the origins of tick toxicoses. Mol Biol. Evol 20:1158-67.

Mans BJ, Neitz AW, 2004a. Adaptation of ticks to a blood-feeding environment: evolution from a functional perspective. Insect Biochem. Mol. Biol. 34:1-17.

Mans BJ, Neitz AW, 2004b. Exon-intron structure of outlier tick lipocalins indicate a monophyletic origin within the larger lipocalin family. Insect Biochem. Mol. Biol. 34:585-94.

Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH, 2002. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res 30:281-3.

Mazet F, Shimeld SM, 2002. Gene duplication and divergence in the early evolution of vertebrates. Curr Opin Genet Dev 12:393-6.

McIntosh JM, Olivera BM, Cruz LJ, Gray WR, 1984. Gamma-carboxyglutamate in a neuroactive toxin. J Biol. Chem 259:14343-6.

Milleron RS, Ribeiro JM, Elnaime D, Soong L, Lanzaro GC, 2004. Negative effect of antibodies against maxadilan on the fitness of the sand fly vector of American visceral leishmaniasis. Am J.Trop Med Hyg 70:278-85.

Milne TJ, Abbenante G, Tyndall JD, Halliday J, Lewis RJ, 2003. Isolation and characterization of a cone snail protease with homology to CRISP proteins of the pathogenesis-related protein superfamily. J Biol. Chem 278:31105-10.

Mochca-Morales J, Martin BM, Possani LD, 1990. Isolation and characterization of helothermine, a novel toxin from Heloderma horridum horridum (Mexican beaded lizard) venom. Toxicon 28:299-309.

Monteiro RQ, Rezaie AR, Ribeiro JM, Francischetti IM, 2004. Ixolaris: A factor Xa heparin binding exosite inhibitor. Biochem. J.

Nakagawa TY, Rudensky AY, 1999. The role of lysosomal proteinases in MHC class II-mediated antigen processing and presentation. Immunol Rev 172:121-9.

Narasimhan S, Koski RA, Beaulieu B, Anderson JF, Ramamoorthi N, Kantor F, Cappello M, Fikrig E, 2002. A novel family of anticoagulants from the saliva of *Ixodes scapularis*. Insect Mol Biol 11:641-50.

Nathoo AN, Moeller RA, Westlund BA, Hart AC, 2001. Identification of neuropeptide-like protein gene families in Caenorhabditiselegans and other species. Proc. Natl. Acad. Sci. U S A 98:14000-5.

Nene V, Lee D, Kang'a S, Skilton R, Shah T, de Villiers E, Mwaura S, Taylor D, Quackenbush J, Bishop R, 2004. Genes transcribed in the salivary glands of female *Rhipicephalus appendiculatus* ticks infected with *Theileria parva*. Insect Biochem. Mol. Biol. 34:1117-28.

Nicholas HB, Jr., Chan SS, Rosenquist GL, 1999. Reevaluation of the determinants of tyrosine sulfation. Endocrine 11:285-92.

Niehrs C, Huttner WB, Carvallo D, Degryse E, 1990. Conversion of recombinant hirudin to the natural form by in vitro tyrosine sulfation. Differential substrate specificities of leech and bovine tyrosylprotein sulfotransferases. J Biol. Chem 265:9314-8.

Nielsen H, Engelbrecht J, Brunak S, von Heijne G, 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. 10:1-6.

Niewiraowski S, McLane MA, Kloczewiak M, Stewart GJ, 1994. Disintegrins and other naturally occurring antagonists of platelet fibrinogen receptors. Sem. Hematol. 31:289-300.

Ohno S, 1970. Evolution by gene duplication. Berlin-Heidelberg-New York: Springer-Verlag.

Ohno S, 1999. Gene duplication and the uniqueness of vertebrate genomes circa 1970-1999. Semin Cell Dev Biol 10:517-22.

Oliver JH, Jr., 1977. Cytogenetics of mites and ticks. Annu Rev Entomol 22:407-29.

Oliver JH, Jr., Nelson BC, 1967. Mite chromosomes: an exceptionally small number. Nature 214:809.

O'Neill SL, Kittayapong P, Braig HR, Andreadis TG, Gonzalez JP, Tesh RB, 1995. Insect densoviruses may be widespread in mosquito cell lines. J Gen Virol 76 (Pt 8):2067-74.

Packila M, Guilfoile PG, 2002. Mating, male Ixodes scapularis express several genes including those with sequence similarity to immunoglobulin-binding proteins and metalloproteases. Exp Appl Acarol 27:151-60.

Paesen GC, Adams PL, Harlos K, Nuttall PA, Stuart DI, 1999. Tick histamine-binding proteins: isolation, cloning, and three- dimensional structure. Mol Cell 3:661-71.

Panzera F, Perez R, Panzera Y, Alvarez F, Scvortzoff E, Salvatella R, 1995. Karyotype evolution in holocentric chromosomes of three related species of triatomines (Hemiptera-Reduviidae). Chromosome Res 3:143-50.

Pittman DD, Wang JH, Kaufman RJ, 1992. Identification and functional importance of tyrosine sulfate residues within recombinant factor VIII. Biochemistry 31:3315-25.

Porter S, Clark IM, Kevorkian L, Edwards DR, 2005. The ADAMTS metalloproteinases. Biochem. J. 386:15-27.

Price PA, Fraser JD, Metz-Virca G, 1987. Molecular cloning of matrix Gla protein: implications for substrate recognition by the vitamin K-dependent gamma-carboxylase. Proc. Natl. Acad. Sci. U S A 84:8335-9.

Rampias TN, Sideris DC, Fragoulis EG, 2003. Cc RNase: the *Ceratitis capitata* ortholog of a novel highly conserved protein family in metazoans. Nucleic Acids Res 31:3092-100.

Rhoads RE, Udenfriend S, 1969. Substrate specificity of collagen proline hydroxylase: hydroxylation of a specific proline residue in bradykinin. Arch Biochem Biophys 133:108-11.

Ribeiro JM, Andersen J, Silva-Neto MA, Pham VM, Garfield MK, Valenzuela JG, 2004. Exploring the sialome of the blood-sucking bug *Rhodnius prolixus*. Insect Biochem. Mol. Biol. 34:61-79.

Ribeiro JM, Francischetti IM, 2003. Role of arthropod saliva in blood feeding: sialome and post-sialome perspectives. Annu. Rev. Entomol. 48:73-88.

Ribeiro JM, Mather TN, 1998. *Ixodes scapularis*: salivary kininase activity is a metallo dipeptidyl carboxypeptidase. Exp Parasitol 89:213-21.

Ribeiro JMC, 1987. *Ixodes dammini:* Salivary anticomplement activity. Exp. Parasitol. 64:347-353.

Ribeiro JMC, 1989. Role of saliva in tick/host associations. Exp. Appl. Acarol. 7:15-20.

Ribeiro JMC, 1995. Blood-feeding arthropods: Live syringes or invertebrate pharmacologists? Infect. Agents Dis. 4:143-152.

Ribeiro JMC, Makoul G, Levine J, Robinson D, Spielman A, 1985. Antihemostatic, antiinflammatory and immunosuppressive properties of the saliva of a tick, *Ixodes dammini*. J. Exp. Med. 161:332-44.

Ribeiro JMC, Makoul G, Robinson D, 1988. *Ixodes dammini*: Evidence for salivary prostacyclin secretion. J. Parasitol. 74:1068-1069.

Ribeiro JMC, Weis JJ, Telford SR, III, 1990. Saliva of the tick *Ixodes dammini* inhibits neutrophil function. Exp. Parasitol. 70:382-8.

Sakamoto N, Mukae H, Fujii T, Ishii H, Yoshioka S, Kakugawa T, Sugiyama K, Mizuta Y, Kadota J, Nakazato M, Kohno S, 2005. Differential effects of alpha- and beta-defensin on cytokine production by cultured human bronchial epithelial cells. Am J. Physiol. Lung. Cell Mol. Physiol. 288:L508-13.

Sangamnatdej S, Paesen GC, Slovak M, Nuttall PA, 2002. A high affinity serotonin- and histamine-binding lipocalin from tick saliva. Insect Mol. Biol. 11:79-86.

Sankoff D, 2001. Gene and genome duplication. Curr Opin Genet Dev 11:681-4.

Scarborough RM, Naughton MA, Teng W, Rose JW, Phillips DR, Nannizzi L, Arfsten A, Campbell AM, Charo IF, 1993. Design of potent and specific integrin antagonists. Peptide antagonists with high specificity for glycoprotein IIb-IIIa. J Biol. Chem 268:1066-73.

Schultz J, Copley RR, Doerks T, Ponting CP, Bork P, 2000. SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res. 28:231-4.

Shao R, Aoki Y, Mitani H, Tabuchi N, Barker SC, Fukunaga M, 2004. The mitochondrial genomes of soft ticks have an arrangement of genes that has remained unchanged for over 400 million years. Insect Mol Biol 13:219-24.

Spielman A, Wilson ML, Levine JF, Piesman J, 1985. Ecology of Ixodes dammini borne human babesiosis and Lyme disease. Ann. Rev. Entomol. 30:439-460.

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA, 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4:41.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG, 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876-82.

Torres AM, Kuchel PW, 2004. The beta-defensin-fold family of polypeptides. Toxicon 44:581-8.

Trager W, 1939. Acquired immunity to ticks. J. Parasitol. 25:57-81.

Ullmann AJ, Lima CM, Guerrero FD, Piesman J, Black WCt, 2005. Genome size and organization in the blacklegged tick, Ixodes scapularis and the Southern cattle tick, Boophilus microplus. Insect Mol Biol 14:217-22.

Valenzuela JG, Charlab R, Mather TN, Ribeiro JM, 2000. Purification, cloning, and expression of a novel salivary anticomplement protein from the tick, *Ixodes scapularis*. J. Biol. Chem. 275:18717-23.

Valenzuela JG, Francischetti IM, Pham VM, Garfield MK, Ribeiro JM, 2003. Exploring the salivary gland transcriptome and proteome of the *Anopheles stephensi* mosquito. Insect Biochem. Mol. Biol. 33:717-32.

Valenzuela JG, Francischetti IMB, Pham VM, Garfield MK, Mather TN, Ribeiro JMC, 2002. Exploring the sialome of the tick, *Ixodes scapularis*. J. Exp. Biol. 205:2843-64.

Vray B, Hartmann S, Hoebeke J, 2002. Immunomodulatory properties of cystatins. Cell Mol Life Sci 59:1503-12.

Wikel SK, 1999. Tick modulation of host immunity: an important factor in pathogen transmission. Int J Parasitol 29:851-9.

Wikel SK, Alarcon-Chaidez FJ, 2001. Progress toward molecular characterization of ectoparasite modulation of host immunity. Vet Parasitol 101:275-87.

Wysoki M, Swirski E, 1968. Karyotypes and sex determination of ten species of phytoseiid mites (Acarina: Mesostigmat). Genetica 39:220-8.

Yamazaki Y, Koike H, Sugiyama Y, Motoyoshi K, Wada T, Hishinuma S, Mita M, Morita T, 2002. Cloning and characterization of novel snake venom proteins that block smooth muscle contraction. Eur J Biochem 269:2708-15.

Zhang B, Ginsburg D, 2004. Familial multiple coagulation factor deficiencies: new biologic insight from rare genetic bleeding disorders. J Thromb Haemost 2:1564-72.

Zhu K, Bowman AS, Dillwith JW, Sauer JR, 1998. Phospholipase A2 activity in salivary glands and saliva of the lone star tick (Acari: Ixodidae) during tick feeding. J. Med. Entomol. 35:500-4.